

# Hierarchies of Local-Optimality Characterizations in Decoding Tanner Codes

Nissim Halabi <sup>\*</sup>      Guy Even <sup>†</sup>

## Abstract

Recent developments in decoding Tanner codes with maximum-likelihood certificates are based on a sufficient condition called local-optimality. We define hierarchies of locally optimal codewords with respect to two parameters. One parameter is related to the distance of the local codes in Tanner codes. The second parameter is related to the finite number of iterations used in iterative decoding. We show that these hierarchies satisfy inclusion properties as these parameters are increased. In particular, this implies that a codeword that is decoded with a certificate using an iterative decoder after  $h$  iterations is decoded with a certificate after  $k \cdot h$  iterations, for every integer  $k$ .

## 1 Introduction

Local optimality is often used as a sufficient condition for successful decoding of finite length codes (see e.g., [WJW05, ADS09]). In this work we focus on two parameters of the local-optimality characterization for Tanner codes [HE11]. The first parameter is related to the distance of the local codes in (expander) Tanner codes. The second parameter is related to the finite number of iterations used in iterative decoding even beyond the girth. We define hierarchies of local optimality with respect to these parameters. These hierarchies provide a partial explanation of two questions about successful decoding with ML-certificates: (1) What is the effect of increasing the distance of the local codes in Tanner codes? (2) What is the effect of increasing the number of iterations beyond the girth in iterative decoding?

*Previous Work:* Density Evolution (DE) is used to study the asymptotic performance of decoding algorithms based on Belief-Propagation (BP) (see e.g., [RU01, CF02]). Convergence of BP-based decoding algorithms was studied in [FK00, WF01, WJW05, JP11]. Note that convergence guaranties do not imply successful decoding after a finite number of iterations. Korada and Urbanke [KU11] provide an asymptotic analysis of iterative decoding “beyond” the girth. Specifically, they prove that one may exchange the order of the limits in DE-analysis of BP-decoding under certain conditions (i.e., variable node degree at least 5 and bounded LLRs). On the other hand, our work focuses on iterative decoding of finite length codes using a finite number of iterations.

---

<sup>\*</sup>School of Electrical Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel.  
E-mail: nissimh@eng.tau.ac.il.

<sup>†</sup>School of Electrical Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel.  
E-mail: guy@eng.tau.ac.il.

Suboptimal decoding of expander Tanner codes was analyzed in many works (see [SS96, BZ04, FS05]). The results in these analyses rely on: (i) the expansion properties of the Tanner graph, and (ii) constant relative distances of the local codes. The error-correcting guaranties in these analyses improve as the relative distance increases.

A new local-optimality characterization for a codeword in a Tanner code w.r.t. any MBIOS channel was presented in [HE11]. A locally-optimal codeword is guaranteed to be both the unique maximum-likelihood (ML) codeword as well as the unique LP-decoding codeword. The characterization of local-optimality for Tanner codes has three parameters: (i) a height  $h \in \mathbb{N}$ , (ii) level weights  $w \in \mathbb{R}_+^h$ , and (iii) a degree  $2 \leq d \leq d^*$ , where  $d^*$  is the minimum local distance.

A new message-passing decoding algorithm, called *normalized weighted min-sum* (NWMS), was presented for Tanner codes with single parity-check (SPC) local-codes [HE11]. The NWMS decoder is guaranteed to compute the ML-codeword in  $h$  iterations provided that a locally-optimal codeword with height  $h$  exists. The number of iterations  $h$  may exceed the girth of the Tanner graph.

*Contribution:* We present a variation of local-optimality called *strong local-optimality*. We prove that if a codeword is strongly locally-optimal, then it is also locally-optimal. Hence, previous results proved for local-optimality [HE11] hold also for strong local-optimality.

We present two hierarchies: (1) A *hierarchy of local-optimality based on degrees*. The degree hierarchy states that a locally optimal codeword  $x$  with degree parameter  $d$  is also locally-optimal with respect to any degree parameter  $d' > d$ . The degree hierarchy implies that the occurrence of local-optimality does not decrease as the degree parameter increases. (2) A *hierarchy of strong local-optimality based on height*. The height hierarchy states that if a codeword  $x$  is strongly locally-optimal with respect to height parameter  $h$ , then it is also strongly locally-optimal with respect to every height parameter that is an integer multiple of  $h$ . The height hierarchy proves, for example, that the performance of iterative decoding with an ML-certificate (e.g., NWMS) of finite-length Tanner codes with SPC local-codes does not degrade as the number of iterations grows, even beyond the girth of the Tanner graph.

**Organization.** In Section 3 we introduce a key procedure used in the proof of the presented hierarchies. In Section 4 we prove that the degree-based hierarchy induces a chain of inclusions of locally optimal codewords and LLRs. In Section 5 we prove a height-based hierarchy over strong local-optimality. We show that strong local-optimality implies local-optimality. Numerical results of strong local-optimality and local-optimality with respect to the height hierarchy are presented in Section 6. We conclude with a discussion in Section 7.

## 2 Preliminaries

**Graph Terminology.** Let  $G = (V, E)$  denote an undirected graph. Let  $\mathcal{N}_G(v)$  denote the set of neighbors of node  $v \in V$ , and let  $\deg_G(v) \triangleq |\mathcal{N}_G(v)|$  denote the degree of node  $v$  in graph  $G$ . A path  $p = (v, \dots, u)$  in  $G$  is a sequence of vertices such that there exists an edge between every two consecutive nodes in the sequence  $p$ . A path  $p$  is *backtrackless* if every two consecutive edges along  $p$  do not close a cycle. Let  $|p|$  denote the number of edges in  $p$ . Let  $\text{girth}(G)$  denote the length of the shortest cycle in  $G$ .

**Tanner-codes.** Let  $G = (\mathcal{V} \cup \mathcal{J}, E)$  denote an edge-labeled bipartite-graph, where  $\mathcal{V} = \{v_1, \dots, v_N\}$  is a set of  $N$  vertices called *variable nodes*, and  $\mathcal{J} = \{C_1, \dots, C_J\}$  is a set of  $J$  vertices called *local-code nodes*. We associate with each local-code node  $C_j$  a linear code  $\bar{\mathcal{C}}^j$  of length  $\deg_G(C_j)$ . Let  $\bar{\mathcal{C}}^\mathcal{J} \triangleq \{\bar{\mathcal{C}}^j : 1 \leq j \leq J\}$  denote the set of *local-codes*, one for each local code node. We say that  $v_i$  *participates* in  $\bar{\mathcal{C}}^j$  if  $(v_i, C_j)$  is an edge in  $E$ .

A word  $x = (x_1, \dots, x_N) \in \{0, 1\}^N$  as an assignment to variable nodes in  $\mathcal{V}$  where  $x_i$  is assigned to  $v_i$ . The *Tanner code*  $\mathcal{C}(G, \bar{\mathcal{C}}^\mathcal{J})$  based on the labeled *Tanner graph*  $G$  is the set of vectors  $x \in \{0, 1\}^N$  such that the projection of  $x$  onto entries associated with  $\mathcal{N}_G(C_j)$  is a codeword in  $\bar{\mathcal{C}}^j$  for every  $j \in \{1, \dots, J\}$ . Let  $d_j$  denote the minimum distance of the local code  $\bar{\mathcal{C}}^j$ . The *minimum local distance*  $d^*$  of a Tanner code  $\mathcal{C}(G, \bar{\mathcal{C}}^\mathcal{J})$  is defined by  $d^* \triangleq \min_j d_j$ . We assume that  $d^* \geq 2$ .

If the bipartite graph is  $(d_L, d_R)$ -regular, then the graph defines a  $(d_L, d_R)$ -regular *Tanner code*. If the Tanner graph is sparse, i.e.,  $|E| = O(N)$ , then it defines a *low-density Tanner code*. Tanner codes with single parity check (SPC) local-codes that are based on sparse Tanner graphs are called *low-density parity-check (LDPC) codes*.

**Communicating over memoryless channels.** Let  $c_i \in \{0, 1\}$  denote the  $i$ th transmitted binary symbol (channel input), and let  $y_i \in \mathbb{R}$  denote the  $i$ th received symbol (channel output). A *memoryless binary-input output-symmetric (MBIOS)* channel is defined by a conditional probability density function  $f(y_i | c_i = a)$  for  $a \in \{0, 1\}$ , that satisfies  $f(y_i | 0) = f(-y_i | 1)$ . In MBIOS channels, the *log-likelihood ratio (LLR)* vector  $\lambda \in \mathbb{R}^N$  is defined by  $\lambda_i(y_i) \triangleq \ln \left( \frac{f(y_i | c_i = 0)}{f(y_i | c_i = 1)} \right)$  for every input bit  $i$ . For a code  $\mathcal{C}$ , *Maximum-Likelihood (ML) decoding* is equivalent to  $\hat{x}^{\text{ML}}(y) = \arg \min_{x \in \mathcal{C}} \langle \lambda(y), x \rangle$ .

**Local-Optimality Characterization.** A new characterization for local-optimality of Tanner codes was presented in [HE11] as extension to [ADS09, Von10]. Local-optimality is defined in Definition 4.

**Definition 1 (Path-Prefix Tree).** Consider a graph  $G = (V, E)$  and a node  $r \in V$ . Let  $\hat{V}$  denote the set of all backtrackingless paths in  $G$  with length at most  $h$  that start at node  $r$ , and let  $\hat{E} \triangleq \{(p_1, p_2) \in \hat{V} \times \hat{V} \mid p_1 \text{ is a prefix of } p_2, |p_1| + 1 = |p_2|\}$ . We identify the empty path in  $\hat{V}$  with  $(r)$ . Denote by  $\mathcal{T}_r^h(G) \triangleq (\hat{V}, \hat{E})$  the path-prefix tree of  $G$  rooted at node  $r$  with height  $h$ .

Path prefix trees of  $G$  that are rooted in variable nodes are often called *computation trees*.

We use the following notation. Because vertices in  $\mathcal{T}_r^h(G)$  are paths in  $G$ , we denote vertices in path-prefix trees by  $p$  and  $q$ . Vertices in  $G$  are denoted by  $u, v, r$ . For a path  $p \in \hat{V}$ , let  $t(p)$  denote the last vertex (*target*) of path  $p$ . Denote by  $\text{Prefix}^+(p)$  the set of proper prefixes of the path  $p$ , i.e.,

$$\text{Prefix}^+(p) = \{q \mid q \text{ is a prefix of } p, 1 \leq |q| < |p|\}.$$

Let  $\mathcal{T}_r^h(G) = (\hat{V}, \hat{E})$  denote a path-prefix tree of a Tanner graph  $G = (\mathcal{V} \cup \mathcal{J}, E)$ . Let  $\hat{\mathcal{V}} \triangleq \{p \mid p \in \hat{V}, t(p) \in \mathcal{V}\}$ , and  $\hat{\mathcal{J}} \triangleq \{p \mid p \in \hat{V}, t(p) \in \mathcal{J}\}$ . Paths in  $\hat{\mathcal{V}}$  are called *variable paths*, and paths in  $\hat{\mathcal{J}}$  are called *local-code paths*.

**Definition 2 ( $d$ -tree).** Denote by  $\mathcal{T}_r^{2h}(G) = (\hat{\mathcal{V}} \cup \hat{\mathcal{J}}, \hat{E})$  the path-prefix tree of a Tanner graph  $G$  rooted at node  $r \in \mathcal{V}$ . A subtree  $\mathcal{T} \subseteq \mathcal{T}_r^{2h}(G)$  is a  $d$ -tree if: (i)  $\mathcal{T}$  is rooted at  $(r)$ , (ii) for

every local-code path  $p \in \mathcal{T} \cap \hat{\mathcal{J}}$ ,  $\deg_{\mathcal{T}}(p) = d$ , and (iii) for every variable path  $p \in \mathcal{T} \cap \hat{\mathcal{V}}$ ,  $\deg_{\mathcal{T}}(p) = \deg_{\mathcal{T}^{2h}}(p)$ .

Let  $\mathcal{T}[r, 2h, d](G)$  denote the set of all  $d$ -trees rooted at  $r$  that are subtrees of  $\mathcal{T}_r^{2h}(G)$ .

**Definition 3** ( $w$ -weighted subtree). *Let  $\mathcal{T} = (\hat{\mathcal{V}} \cup \hat{\mathcal{J}}, \hat{E})$  denote a subtree of  $\mathcal{T}_r^{2h}(G)$ , and let  $w = (w_1, \dots, w_h) \in \mathbb{R}_+^h \setminus \{0^h\}$  denote a non-negative weight vector. Let  $w_{\mathcal{T}} : \hat{\mathcal{V}} \rightarrow \mathbb{R}$  denote a weight function based on weight vector  $w$  for variable paths  $p \in \hat{\mathcal{V}}$  defined as follows. If  $p$  is an empty variable path, then  $w_{\mathcal{T}}(p) = 0$ . Otherwise,*

$$w_{\mathcal{T}}(p) \triangleq \frac{w_{\ell}}{\|w\|_1} \cdot \frac{1}{\deg_G(t(p))} \cdot \prod_{q \in \text{Prefix}^+(p)} \frac{1}{\deg_{\mathcal{T}}(q) - 1}, \quad (1)$$

where  $\ell = \lceil \frac{|p|}{2} \rceil$ . We refer to  $w_{\mathcal{T}}$  as a  $w$ -weighted subtree.

For any  $w$ -weighted subtree  $w_{\mathcal{T}}$  of  $\mathcal{T}_r^{2h}(G)$ , let  $\pi_{G, \mathcal{T}, w} : \mathcal{V} \rightarrow \mathbb{R}$  denote a function whose values correspond to the projection of  $w_{\mathcal{T}}$  to the Tanner graph  $G$ . That is, for every variable node  $v$  in  $G$ ,

$$\pi_{G, \mathcal{T}, w}(v) \triangleq \sum_{\{p \in \mathcal{T} \mid t(p) = v\}} w_{\mathcal{T}}(p). \quad (2)$$

For a Tanner code  $\mathcal{C}(G)$ , let  $\mathcal{B}_d^{(w)} \subseteq [0, 1]^N$  denote the set of all projections of  $w$ -weighted  $d$ -trees to  $G$ . That is,

$$\mathcal{B}_d^{(w)} \triangleq \left\{ \pi_{G, \mathcal{T}, w} \mid \mathcal{T} \in \bigcup_{r \in \mathcal{V}} \mathcal{T}[r, 2h, d](G) \right\}. \quad (3)$$

Vectors in  $\mathcal{B}_d^{(w)}$  are referred to as *deviations*. For two vectors  $x \in \{0, 1\}^N$  and  $f \in [0, 1]^N$ , let  $x \oplus f \in [0, 1]^N$  denote the *relative point* defined by  $(x \oplus f)_i \triangleq |x_i - f_i|$  [Fel03].

**Definition 4** (local-optimality, [HE11]). *A codeword  $x \in \mathcal{C}(G)$  is  $(h, w, d)$ -locally optimal with respect to  $\lambda \in \mathbb{R}^N$  if for all vectors  $\beta \in \mathcal{B}_d^{(w)}$ ,*

$$\langle \lambda, x \oplus \beta \rangle > \langle \lambda, x \rangle. \quad (4)$$

**Theorem 5** (local-optimality is sufficient for ML and LP, [HE11]). *Let  $\lambda \in \mathbb{R}^N$  denote the LLR vector received from the channel. If  $x$  is an  $(h, w, d)$ -locally optimal codeword w.r.t.  $\lambda$  and some  $2 \leq d \leq d^*$ , then (1)  $x$  is the unique maximum-likelihood codeword w.r.t.  $\lambda$ , and (2)  $x$  is the unique optimal solution of the LP-decoder given  $\lambda$ .*

For two vectors  $y, z \in \mathbb{R}^N$ , let “ $*$ ” denote coordinatewise multiplication, i.e.,  $y * z \triangleq (y_1 \cdot z_1, \dots, y_N \cdot z_N)$ .

**Proposition 6** ([HE11]). *For every  $\lambda \in \mathbb{R}^N$  and every  $\beta \in [0, 1]^N$ ,*

$$\langle (-1)^x * \lambda, \beta \rangle = \langle \lambda, x \oplus \beta \rangle - \langle \lambda, x \rangle.$$

The following proposition states that the mapping  $(x, \lambda) \mapsto (0^N, (-1)^x * \lambda)$  preserves local-optimality.

**Proposition 7** (symmetry of local-optimality, [HE11]). *For every  $x \in \mathcal{C}$ ,  $x$  is  $(h, w, d)$ -locally optimal for  $\lambda$  if and only if  $0^N$  is  $(h, w, d)$ -locally optimal for  $(-1)^x * \lambda$ .*

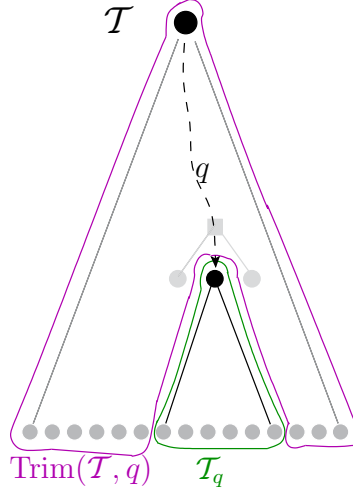


Figure 1: Trimmed tree of  $\mathcal{T}$  induced by  $q$ .

### 3 Trimming Subtrees from a Path-Prefix Tree

Let  $\mathcal{T}_q$  denote the subtree of a path-prefix tree  $\mathcal{T}$  hanging from path  $q$ , i.e., the subtree induced by  $\hat{V}_q \triangleq \{p \in \hat{\mathcal{V}} \cup \hat{\mathcal{J}} \mid q \in \text{Prefix}^+(p) \text{ or } p = q\}$  (see Figure 1). Let  $\text{Trim}(\mathcal{T}, q)$  denote the *trimmed-tree* of  $\mathcal{T}$  induced by  $q$  obtained by deleting the subtree  $\mathcal{T}_q$  from  $\mathcal{T}$ . Formally,  $\text{Trim}(\mathcal{T}, q)$  is the path-prefix subtree of  $\mathcal{T}$  induced by  $\hat{\mathcal{V}} \cup \hat{\mathcal{J}} \setminus \hat{V}_q$ . Note that if  $q'$  is a sibling of  $q$  (i.e.,  $q'$  differs from  $q$  only in the last edge), then the degree of the parent of  $q$  and  $q'$  decreases as a result of trimming  $\hat{V}_q$ . Hence,  $w_{\mathcal{T}}(q'') < w_{\text{Trim}(\mathcal{T}, q)}(q'')$  for every variable path  $q'' \in \hat{V}_{q'}$ .

The proofs of hierarchies presented in the following sections are based on the following lemma.

**Lemma 8.** *Let  $\mathcal{T}$  denote a subtree of a path-prefix tree  $\mathcal{T}_r^{2h}(G)$ . For every path  $p \in \mathcal{T}$  with at least two children in  $\mathcal{T}$ , there exists at least one child  $q$  of  $p$ , such that*

$$\langle \lambda, \pi_{G, \mathcal{T}, w} \rangle \geq \langle \lambda, \pi_{G, \text{Trim}(\mathcal{T}, q), w} \rangle.$$

*Proof.* See Appendix A. □

### 4 Degree Hierarchy of Local-Optimality

Let  $\Lambda \subseteq \mathbb{R}^N$  denote a set of vectors. Denote by  $\text{LO}_{\mathcal{C}, \Lambda}(h, w, d)$  the set of pairs  $(x, \lambda) \in \mathcal{C} \times \Lambda$  such that  $x$  is  $(h, w, d)$ -locally optimal w.r.t.  $\lambda$ . Formally,

$$\text{LO}_{\mathcal{C}, \Lambda}(h, w, d) \triangleq \{(x, \lambda) \in \mathcal{C} \times \Lambda \mid x \text{ is } (h, w, d)\text{-locally optimal w.r.t. } \lambda\}. \quad (5)$$

The following theorem derives an hierarchy on the “density” of deviations in local-optimality characterization.

**Theorem 9** (*d*-Hierarchy of local-optimality). *Let  $2 \leq d < d^*$ . For every  $\Lambda \subseteq \mathbb{R}^N$ ,*

$$\text{LO}_{\mathcal{C}, \Lambda}(h, w, d) \subseteq \text{LO}_{\mathcal{C}, \Lambda}(h, w, d + 1).$$

*Proof.* We prove the contrapositive statement. Assume that  $x$  is not  $(h, w, d+1)$ -locally optimal w.r.t.  $\lambda$ . By Proposition 7,  $0^N$  is not  $(h, w, d+1)$ -locally optimal w.r.t.  $\lambda^0 \triangleq (-1)^x * \lambda$ . Hence, there exists a deviation  $\beta = \pi_{G, \mathcal{T}, w} \in \mathcal{B}_d^{(w)}$  such that  $\langle \lambda^0, \beta \rangle \leq 0$ . Let  $\mathcal{T}$  denote the  $(d+1)$ -tree that corresponds to the deviation  $\beta$ .

Consider the following iterative trimming process. Start with the  $(d+1)$ -tree  $\mathcal{T}$  and let  $\mathcal{T} \leftarrow \mathcal{T}'$ ; While there exists a local-code path  $p \in \mathcal{T}'$  such that  $\deg_{\mathcal{T}'}(p) = d+1$  do:  $\mathcal{T}' \leftarrow \text{Trim}(\mathcal{T}', q)$  where  $q$  is a child of  $p$  such that  $\langle \lambda^0, \pi_{G, \mathcal{T}', w} \rangle \geq \langle \lambda^0, \pi_{G, \text{Trim}(\mathcal{T}', q), w} \rangle$ .

Lemma 8 guarantees that the iterative trimming process halts with a  $d$ -tree  $\mathcal{T}'$  whose corresponding deviation  $\beta' = \pi_{G, \mathcal{T}', w}$  satisfies  $\langle \lambda^0, \beta' \rangle \leq \langle \lambda^0, \beta \rangle \leq 0$ . We conclude by Proposition 7 that  $x$  is not  $(h, w, d)$ -locally optimal w.r.t.  $\lambda$ , as required.  $\square$

We therefore have for every  $2 \leq d < d^*$ ,

$$\Pr_\lambda \{x \text{ is } (h, w, d+1)\text{-locally optimal w.r.t. } \lambda\} \geq \Pr_\lambda \{x \text{ is } (h, w, d)\text{-locally optimal w.r.t. } \lambda\}.$$

## 5 Height Hierarchy of Strong Local-Optimality

In this section we introduce a new combinatorial characterization named *strong local-optimality*. We prove that if a codeword is strongly locally-optimal then it is also locally-optimal. The other direction is not true in general.

**Definition 10** (reduced  $d$ -tree). Denote by  $\mathcal{T}_r^{2h}(G) = (\hat{\mathcal{V}} \cup \hat{\mathcal{J}}, \hat{E})$  the path-prefix tree of a Tanner graph  $G$  rooted at node  $r \in \mathcal{V}$ . A subtree  $\mathcal{T} \subseteq \mathcal{T}_r^{2h}(G)$  is a reduced  $d$ -tree if: (i)  $\mathcal{T}$  is rooted at  $r$ , (ii)  $\deg_{\mathcal{T}}((r)) = \deg_G(r) - 1$ , (iii) for every local-code path  $p \in \mathcal{T} \cap \hat{\mathcal{J}}$ ,  $\deg_{\mathcal{T}}(p) = d$ , and (iv) for every non-empty variable path  $p \in \mathcal{T} \cap \hat{\mathcal{V}}$ ,  $\deg_{\mathcal{T}}(p) = \deg_{\mathcal{T}_r^{2h}}(p)$ .

The only difference between Definition 2 ( $d$ -tree) to a reduced  $d$ -tree is that the degree of the root in a reduced  $d$ -tree is smaller by 1 (as if the root itself hangs from an edge).

Let  $\mathcal{T}^{\text{red}}[r, 2h, d](G)$  denote the set of all reduced  $d$ -trees rooted at  $r$  that are subtrees of  $\mathcal{T}_r^{2h}(G)$ . For a Tanner code  $\mathcal{C}(G)$ , let  $\overline{\mathcal{B}}_d^{(w)} \subseteq [0, 1]^N$  denote the set of all projections of  $w$ -weighted reduced  $d$ -trees to  $G$ . That is,

$$\overline{\mathcal{B}}_d^{(w)} \triangleq \{\pi_{G, \mathcal{T}, w} \mid \mathcal{T} \in \bigcup_{r \in \mathcal{V}} \mathcal{T}^{\text{red}}[r, 2h, d](G)\}. \quad (6)$$

Vectors in  $\overline{\mathcal{B}}_d^{(w)}$  are referred to as *reduced deviations*.

The following definition is analogues to Definition 4 (local-optimality) using reduced deviations instead of deviations.

**Definition 11** (strong local-optimality). Let  $\mathcal{C}(G) \subset \{0, 1\}^N$  denote a Tanner code with minimum local distance  $d^*$ . Let  $w \in \mathbb{R}_+^h \setminus \{0^h\}$  denote a non-negative weight vector of length  $h$  and let  $2 \leq d \leq d^*$ . A codeword  $x \in \mathcal{C}(G)$  is  $(h, w, d)$ -strong locally-optimal with respect to  $\lambda \in \mathbb{R}^N$  if for all vectors  $\beta \in \overline{\mathcal{B}}_d^{(w)}$ ,

$$\langle \lambda, x \oplus \beta \rangle > \langle \lambda, x \rangle. \quad (7)$$

Denote by  $\text{SLO}_{\mathcal{C},\Lambda}(h, w, d)$  the set pairs  $(x, \lambda) \in \mathcal{C} \times \Lambda$  such that  $x$  is  $(h, w, d)$ -strong locally-optimal w.r.t.  $\lambda$ . Formally,

$$\text{SLO}_{\mathcal{C},\Lambda}(h, w, d) \triangleq \{(x, \lambda) \in \mathcal{C} \times \Lambda \mid x \text{ is } (h, w, d)\text{-strong locally-optimal w.r.t. } \lambda\}. \quad (8)$$

The following lemma states that if a codeword  $x$  is strongly locally-optimal w.r.t.  $\lambda$ , then  $x$  is locally-optimal w.r.t.  $\lambda$ .

**Lemma 12.** *For every  $\Lambda \subseteq \mathbb{R}^N$ ,*

$$\text{SLO}_{\mathcal{C},\Lambda}(h, w, d) \subseteq \text{LO}_{\mathcal{C},\Lambda}(h, w, d).$$

*Proof.* We prove the contrapositive statement. Assume that  $x$  is not  $(h, w, d)$ -locally optimal w.r.t.  $\lambda$ . By Proposition 7,  $0^N$  is not  $(h, w, d)$ -locally optimal w.r.t.  $\lambda^0 \triangleq (-1)^x * \lambda$ . Hence, there exists a deviation  $\beta = \pi_{G,\mathcal{T},w} \in \mathcal{B}_d^{(w)}$  such that  $\langle \lambda^0, \beta \rangle \leq 0$ . Let  $\mathcal{T}$  denote the  $d$ -tree that corresponds to the deviation  $\beta$ .

Denote by  $(r)$  the root of  $\mathcal{T}$ . By Lemma 8, the root  $(r)$  has a child  $q$  such that  $\langle \lambda^0, \pi_{G,\mathcal{T},w} \rangle \geq \langle \lambda^0, \pi_{G,\text{Trim}(\mathcal{T},q),w} \rangle$ . Note that  $\text{Trim}(\mathcal{T}, q)$  is a reduced  $d$ -tree rooted at  $r$ . Moreover, the corresponding reduced deviation  $\beta' = \pi_{G,\mathcal{T}',w}$  satisfies  $\langle \lambda^0, \beta' \rangle \leq \langle \lambda^0, \beta \rangle \leq 0$ . We conclude by Proposition 7 that  $x$  is not  $(h, w, d)$ -strong locally-optimal w.r.t.  $\lambda$ , as required.  $\square$

Following Lemma 12 and Theorem 5 we have the following corollary.

**Corollary 13** (strong local-optimality is sufficient for both ML and LP). *Let  $\mathcal{C}(G)$  denote a Tanner code with minimum local distance  $d^*$ . Let  $h \in \mathbb{N}_+$  and  $w \in \mathbb{R}_+^h$ . Let  $\lambda \in \mathbb{R}^N$  denote the LLR vector received from the channel. If  $x$  is an  $(h, w, d)$ -strong locally-optimal codeword w.r.t.  $\lambda$  and some  $2 \leq d \leq d^*$ , then (1)  $x$  is the unique maximum-likelihood codeword w.r.t.  $\lambda$ , and (2)  $x$  is the unique solution of LP-decoding given  $\lambda$ .*

Consider a weight vector  $\bar{w} \in \mathbb{R}^{k \cdot h}$ , and let  $\bar{w} = \bar{w}^1 \circ \bar{w}^2 \circ \dots \circ \bar{w}^k$  denote its decomposition to  $k$  weight vectors  $\bar{w}^i \in \mathbb{R}^h$ .  $\bar{w} \in \mathbb{R}^{k \cdot h}$  is a  $k$ -legal extension of  $w \in \mathbb{R}^h$  if  $\exists \alpha \in \mathbb{R}^k$  such that  $\bar{w}^i = \alpha_i \cdot w$ . Note that if  $\bar{w} \in \mathbb{R}^{k \cdot h}$  is geometric, then it is a  $k$ -legal extension of  $\bar{w}^1$  in its decomposition.

The following theorem derives an hierarchy on the height of reduced deviations of strong local-optimality characterization.

**Theorem 14** ( $h$ -Hierarchy of strong LO). *For every  $\Lambda \subseteq \mathbb{R}^N$ , if  $\bar{w} \in \mathbb{R}^{k \cdot h}$  is a  $k$ -legal extension of  $w \in \mathbb{R}^h$ , then*

$$\text{SLO}_{\mathcal{C},\Lambda}(h, w, d) \subseteq \text{SLO}_{\mathcal{C},\Lambda}(k \cdot h, \bar{w}, d).$$

*Proof.* We prove the contrapositive statement. Assume that  $x$  is not  $(k \cdot h, \bar{w}, d)$ -strong locally-optimal w.r.t.  $\lambda$ . Proposition 6 implies that  $0^N$  is not  $(k \cdot h, \bar{w}, d)$ -strong locally-optimal w.r.t.  $\lambda^0 \triangleq (-1)^x * \lambda$ . Hence, there exists a reduced deviation  $\beta = \pi_{G,\mathcal{T},\bar{w}} \in \mathcal{B}_d^{(\bar{w})}$  such that  $\langle \lambda^0, \beta \rangle \leq 0$ . Let  $\mathcal{T}$  denote the reduced  $d$ -tree that corresponds to the reduced deviation  $\beta$ .

Let  $\{\mathcal{T}^i\}$  denote a decomposition of  $\mathcal{T}$  to reduced  $d$ -trees of height  $2h$  as shown in Figure 2, where leaves of a subtree are the roots of other subtrees. Let  $p^i$  denote the root of a reduced  $d$ -tree  $\mathcal{T}^i$  in the decomposition of  $\mathcal{T}$ . Let  $\text{order}(\mathcal{T}^i) \triangleq \lfloor |p^i|/h \rfloor$ . Namely, the order of  $\mathcal{T}^i$  equals to its level in the decomposition. Note that

$$\pi_{G,\mathcal{T},\bar{w}} = \sum_{\{\mathcal{T}^i\}} \alpha_{\text{order}(\mathcal{T}^i)} \cdot \pi_{G,\mathcal{T}^i,w}.$$

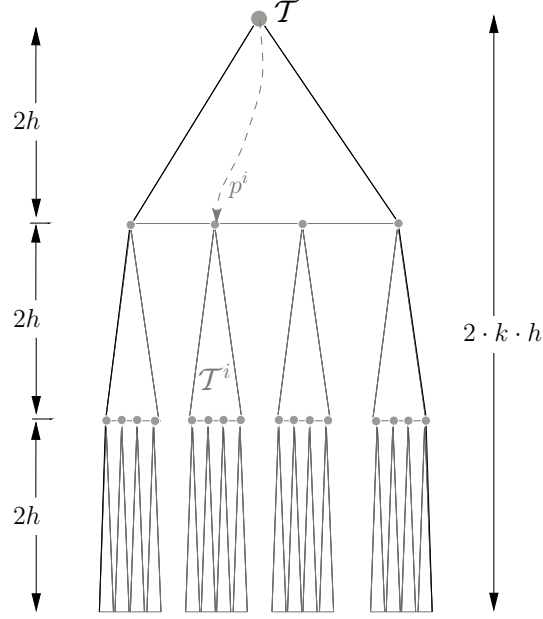


Figure 2: Decomposition of a reduced  $d$ -tree  $\mathcal{T}$  of height  $2kh$  to a set of subtrees  $\{\mathcal{T}^i\}$  that are reduced  $d$ -trees of height  $2h$ .

Because  $\langle \lambda^0, \beta \rangle \leq 0$ , we conclude by averaging that there exists at least one reduced  $d$ -tree  $\mathcal{T}^* \in \{\mathcal{T}^i\}$  of height  $2h$  such that  $\langle \lambda^0, \pi_{G, \mathcal{T}^*, w} \rangle \leq 0$ . Hence,  $0^N$  is not  $(h, w, d)$ -strong locally-optimal w.r.t.  $\lambda^0$ . We apply Proposition 6 again, and conclude that  $x$  is not  $(h, w, d)$ -strong locally-optimal w.r.t.  $\lambda$ , as required.  $\square$

## 6 Numerical Results

We conducted simulations to demonstrate two phenomena. First, we checked the gap between strong local optimality and local optimality. Second, we checked the effect of increasing the number of iterations on successful decoding with ML-certificates.

We chose a  $(3, 6)$ -regular LDPC code with blocklength  $N = 1008$  and girth  $g = 6$  [Mac]. We simulated a set  $\Lambda_p$  of 5000 LLR vectors corresponding to the all zeros codeword with respect to a BSC with crossover probability  $p \in \{0.04, 0.05, 0.06\}$ . We used unit level weights, i.e.,  $w = 1^h$ .

Let  $\text{SLO}_{0^N, \Lambda_p}(h, w, 2)$  (resp.,  $\text{LO}_{0^N, \Lambda_p}(h, w, 2)$ ) denote the set of LLR vectors  $\lambda \in \Lambda_p$  such that  $0^N$  is strongly locally-optimal (resp., locally optimal) w.r.t.  $\lambda$ .

Figure 3 depicts cardinality of  $\text{SLO}_{0^N, \Lambda_p}(h, w, 2)$  and  $\text{LO}_{0^N, \Lambda_p}(h, w, 2)$  as a function of  $h$ , for three values of  $p$ . The results suggest that, in this setting, the sets  $\text{SLO}_{\{0^N\}, \Lambda_p}(h, w, 2)$  and  $\text{LO}_{\{0^N\}, \Lambda_p}(h, w, 2)$  coincide as  $h$  grows. This suggests also that the containment in Lemma 12 is asymptotically tight. That is, for large height  $h$ , strong local optimality is very close to local-optimality.

The results also suggest that the number of iterations needed to obtain reasonable decoding with ML-certificates is far greater than the girth. Clearly, the “tree property” that DE analysis



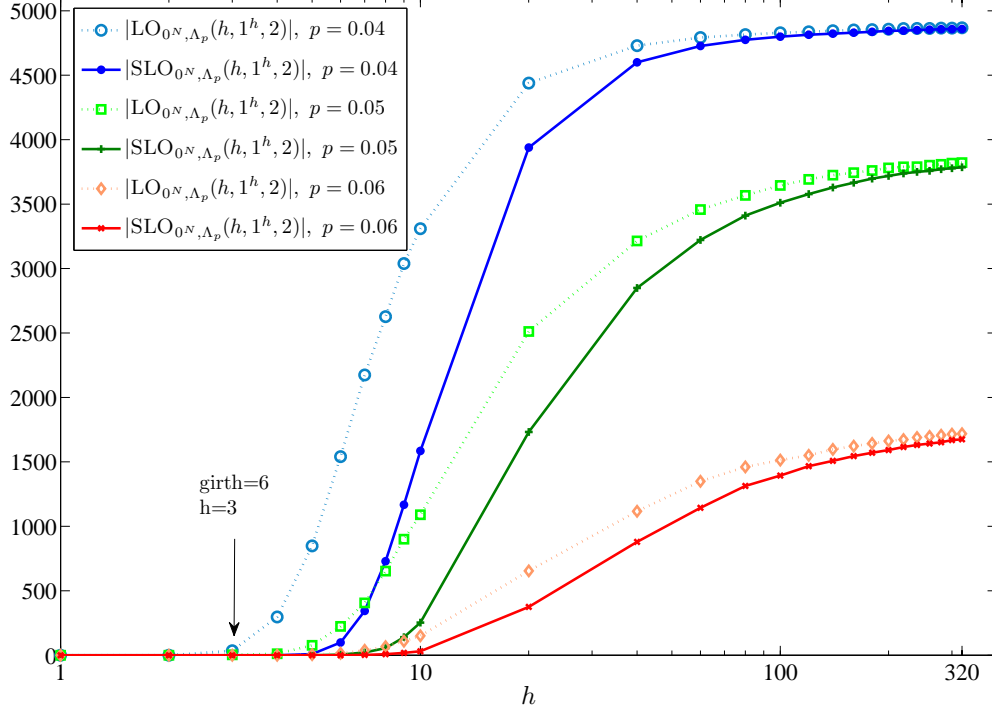


Figure 3: Growth of strong local-optimality and local-optimality as a function of the height  $h$ .  $|\Lambda_p| = 5000$  for  $p \in \{0.04, 0.05, 0.06\}$ .

relies on does not hold for so many iterations. Indeed, the simulated crossover probabilities are in the “waterfall” region of the word error rate curve with respect to NWMS. We are not aware of any analytic explanation of this phenomena in finite length codes.

Another result of the simulation is that  $SLO_{0^N, \Lambda_p}(h, w, 2) \subseteq SLO_{0^N, \Lambda_p}(h+1, w, 2)$ . Namely, once a codeword is strongly locally-optimal for  $\lambda$  with height  $h$ , then it is also strongly locally-optimal for any height  $h' > h$ . This exhibited strengthening of the height hierarchy result is not true in general. Counterexamples can be obtained for other level weights  $w$  and Tanner codes.

## 7 Discussion

The degree hierarchy supports the improvement in the lower bounds for the threshold of the crossover probability  $p$  of a  $BSC_p$  as a function of  $d$  (see [HE11, Theorem 27]). These lower bounds are proved by analyzing the probability of a locally optimal codeword as a function of  $p$  and the degree  $d$ . For example, consider any  $(2, 16)$ -regular Tanner code with minimum local-distance 4 whose Tanner graph has logarithmic girth in the blocklength. The bounds in [HE11] imply a lower bound on the threshold of  $p_0 = 0.019$  with respect to degree  $d = 3$ . On the other hand, the lower bound on the threshold increases to  $p_0 = 0.044$  with respect to degree  $d = 4$ .

The height hierarchy implies that if a codeword  $x$  is  $(h, w, 2)$ -strong locally-optimal w.r.t. an LLR vector  $\lambda$ , then it is also strongly locally-optimal with respect to any legal extension of level weights  $w$  with larger height  $h'$ .

Consider a Tanner code with single parity-check local-codes. Assume that  $x$  is strongly

locally-optimal codeword w.r.t.  $\lambda$  based on a height parameter  $h$ . Because strong local-optimality implies local-optimality, following [HE11, Theorem 16], we conclude that iterative message-passing decoding by NWMS is guaranteed to decode the ML-certified codeword  $x$  after  $k \cdot h$  iterations, for every  $k \in \mathbb{N}_+$ . This gives the following new insight of convergence. If a codeword  $x$  is decoded after  $h$  iterations and is certified to be strongly locally-optimal (and hence ML-optimal), then  $x$  is the outcome of NWMS infinitely many times (i.e., whenever the number of iterations is a multiple of  $h$ ).

## 8 Conclusion

We present hierarchies of local optimality with respect to two parameters of the local-optimality characterization for Tanner codes [HE11]. One hierarchy is based on the local code node degrees in the deviations. We prove containment, namely, the set of locally optimal codewords with respect to degree  $d + 1$  is a superset of the set of locally optimal codewords with respect to degree  $d$ .

The second hierarchy is based on the height of the deviations. We prove that, for geometric level weights, a strongly locally optimal codeword is infinitely often strongly locally optimal. This result implies that a codeword that is decoded with a certificate using the iterative decoder NWMS after  $h$  iterations is decoded with a certificate after  $k \cdot h$  iterations, for every integer  $k$ .

## A Proof of Lemma 8

*Proof.* Consider a path  $p \in \mathcal{T}$ . Then,

$$\langle \lambda, \pi_{G, \mathcal{T}, w} \rangle = \underbrace{\sum_{q \in \hat{\mathcal{V}} \setminus \hat{\mathcal{V}}_p} \lambda_{t(q)} \cdot w_{\mathcal{T}}(q)}_{(a)} + \underbrace{\sum_{q \in \hat{\mathcal{V}} \cap \hat{\mathcal{V}}_p} \lambda_{t(q)} \cdot w_{\mathcal{T}}(q)}_{(b)}. \quad (9)$$

We deal with terms (a) and (b) in Equation (9) separately.

First we deal with term (a). Let  $q' \in \mathcal{N}_{\mathcal{T}}(p)$ ,  $|q'| = |p| + 1$ , denote a child of  $p$ . Because  $p \notin \text{Prefix}^+(q')$  for the paths accumulated in term (a), it holds that

$$\sum_{q \in \hat{\mathcal{V}} \setminus \hat{\mathcal{V}}_p} \lambda_{t(q)} \cdot w_{\mathcal{T}}(q) = \sum_{q \in \hat{\mathcal{V}} \setminus \hat{\mathcal{V}}_p} \lambda_{t(q)} \cdot w_{\text{Trim}(\mathcal{T}, q')}(q) \quad (10)$$

Hence, term (a) remains unchanged under trimming children of  $p$  from  $\mathcal{T}$ .

It remains to show that there exists a child  $q^1$  of  $p$  whose trimming does not increase term (b). Let  $\text{cost}_{\mathcal{T}}(\mathcal{T}_q) \triangleq \sum_{q \in \hat{\mathcal{V}}_q} \lambda_{t(q)} w_{\mathcal{T}}(q)$  denote the cost of  $\mathcal{T}_q$  with respect to  $\mathcal{T}$ . Note that term (b) equals to  $\text{cost}_{\mathcal{T}}(\mathcal{T}_p)$ , and

$$\text{cost}_{\mathcal{T}}(\mathcal{T}_p) = \lambda_{t(p)} w_{\mathcal{T}}(p) + \sum_{\{q \in \mathcal{N}_{\mathcal{T}}(p) : |q|=|p|+1\}} \text{cost}_{\mathcal{T}}(\mathcal{T}_q). \quad (11)$$

Consider two children  $q_1$  and  $q_2$  of  $p$ . By Definition 3, for every variable path  $q \in \mathcal{T}_{q_2}$ ,  $w_{\mathcal{T}}(q) = \frac{(\deg_{\mathcal{T}}(p)-1)}{(\deg_{\mathcal{T}}(p)-2)} w_{\text{Trim}(\mathcal{T}, q_1)}(q)$ . Hence,

$$\text{cost}_{\mathcal{T}}(\mathcal{T}_{q_2}) = \frac{(\deg_{\mathcal{T}}(p)-1)}{(\deg_{\mathcal{T}}(p)-2)} \text{cost}_{\text{Trim}(\mathcal{T}, q_1)}(\mathcal{T}_{q_2}). \quad (12)$$

Let  $q^{\min} \triangleq \arg \min \{\text{cost}_{\mathcal{T}}(\mathcal{T}_q) \mid q \in \mathcal{N}_{\mathcal{T}}(p), |q| = |p| + 1\}$ . Namely,  $q^{\min}$  is a child of  $p$ , for which the subtree hanging from it has a minimum cost. From Equations (11) and (12), it follows by averaging that  $\text{cost}_{\mathcal{T}}(\mathcal{T}_p) \geq \text{cost}_{\text{Trim}(\mathcal{T}, q^{\min})}(\mathcal{T}_p)$ . Hence, trimming the subtree that hangs from  $q^{\min}$  decreases term (b) in Equation (9), and the lemma follows.  $\square$

## References

- [ADS09] S. Arora, C. Daskalakis, and D. Steurer, “Message passing algorithms and improved LP decoding,” in *Proc. 41st Annual ACM Symp. Theory of Computing (STOC’09)*, Bethesda, MD, USA, pp. 3–12, 2009.
- [BZ04] A. Barg and G. Zémor, “Error exponents of expander codes under linear-complexity decoding,” *SIAM J. Discr. Math.*, vol. 17, no. 3, pp 426–445, 2004.
- [CF02] J. Chen and M. P. C. Fossorier, “Density evolution for two improved BP-Based decoding algorithms of LDPC codes,” *IEEE Commun. Lett.*, vol. 6, no. 5, pp. 208–210, May 2002.
- [Fel03] J. Feldman, “Decoding error-correcting codes via linear programming,” Ph.D. dissertation, MIT, Cambridge, MA, 2003.
- [FK00] B.J. Frey and R. Koetter, “Exact inference using the attenuated max-product algorithm,” In *Advanced Mean Field Methods: Theory and Practice*, Cambridge, MA: MIT Press, 2000.
- [FS05] J. Feldman and C. Stein, “LP decoding achieves capacity,” in *Proc. Symp. Discrete Algorithms (SODA’05)*, Vancouver, Canada, Jan. 2005, pp. 460–469.
- [HE11] G. Even and N. Halabi, “On decoding Tanner codes with local-optimality guarantees,” *CoRR*, <http://arxiv.org/abs/1107.2677>, 45 pp., Jul. 2011.
- [JP11] Y.-Y. Jian and H.D. Pfister, “Convergence of weighted min-sum decoding via dynamic programming on coupled trees,” *CoRR*, <http://arxiv.org/abs/1107.3177>, Jul. 2011.
- [KU11] S. B. Korada and R. L. Urbanke, “Exchange of limits: Why iterative decoding works,” *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2169–2187, Apr. 2011.
- [Mac] D. MacKay, *Encyclopedia of Sparse Graph Codes*. Available: <http://www.inference.phy.cam.ac.uk/mackay/codes/>
- [RU01] T. Richardson and R. Urbanke, “The capacity of low-density parity-check codes under message-passing decoding,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.
- [SS96] M. Sipser and D. A. Spielman, “Expander codes,” *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 1710–1722, Nov. 1996.

- [Von10] P. Vontobel, “A factor-graph-based random walk, and its relevance for LP decoding analysis and Bethe entropy characterization,” in *Proc. Information Theory and Applications Workshop*, UC San Diego, LA Jolla, CA, USA, Jan. 31-Feb. 5, 2010.
- [WF01] Y. Weiss and W. T. Freeman, “On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 736–744, Feb. 2001.
- [WJW05] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, “MAP estimation via agreement on trees: message-passing and linear programming,” *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3697–3717, Nov. 2005.